

MONITORING THE QUALITY OF OPERATIONAL AND SEMI-OPERATIONAL SATELLITE PRECIPITATION ESTIMATES - THE IPWG VALIDATION / INTERCOMPARISON STUDY

Elizabeth Ebert

Bureau of Meteorology Research Centre
GPO Box 1289K, Melbourne, Vic., Australia 3001

ABSTRACT

A number of satellite rainfall products are now freely available in near real time on the internet, some of which are so newly developed that a comprehensive evaluation has not yet been published. Since early 2003 we have conducted a daily validation and intercomparison of several operational and semi-operational satellite rainfall algorithms using tropical and mid-latitude gauge and radar rainfall observations over Australia, the US, and western Europe. 24 h forecasts from a small number of numerical weather prediction models are included for comparison. The validation products are updated daily and displayed on the world wide web, and include maps, time series, and statistics. These will help algorithm developers to clarify the strengths and weaknesses of their algorithms, and users of satellite rainfall estimates to better understand the accuracy and limitations of those products.

1. INTRODUCTION

Satellite rainfall estimates from a variety of operational and semi-operational algorithms are easily obtainable via the web or FTP, and are being used for many diverse meteorological, climate, hydrological, agricultural, and other applications. To use these products appropriately it is important to have an idea of their accuracy and expected error characteristics. Different users have different accuracy requirements. For example, hydrologists need to know the accuracy of rain volume estimates, while forecasters may be more interested in the accuracy for heavy rain detection.

During the past 15 years a series of algorithm intercomparison projects have evaluated satellite rainfall estimates from a large number of infrared (IR), microwave (MW), and combined IR-MW algorithms. The WetNet Precipitation Intercomparison Projects (PIP-1, -2, -3), and the Global Precipitation Climatology Project (GPCP) Algorithm Intercomparison Projects (AIP-1, -2, -3) assessed the accuracy of satellite rainfall estimates of the occurrence and amount of rainfall on instantaneous, daily, and monthly time scales, and on spatial scales of individual pixels to 2.5° latitude/longitude. Briefly, these studies found that for instantaneous rainfall, algorithms using MW observations gave more accurate estimates than those using IR observations only. As the time scale increased to daily and monthly, the IR and MW-IR algorithms performed comparably or even better than the MW-only algorithms. This was mainly due to the much greater sampling from the geostationary satellites carrying IR instruments, compared to the limited sampling from microwave instruments on polar orbiting satellites. Furthermore, greater accuracy was shown for tropical,

convective, and summertime rainfall, and poorer accuracy for stratiform, mid-latitude, and wintertime rainfall.

From the late 1980's when the algorithm intercomparison programs began till the late 1990's when they concluded, the rainfall community saw an overall improvement in the accuracy of the satellite estimates. This was due in no small part to the lessons learned from the AIPs and PIPs. Since the early 2000's new algorithms have been developed, notably a class of algorithms that use IR imagery to "advect" and "evolve" the rainfall between successive microwave images. These newer algorithms need to be evaluated alongside existing satellite algorithms.

In 2003 the International Precipitation Working Group began a project to validate and intercompare several operational and semi-operational satellite rainfall estimates. Satellite estimates of 24 h accumulated rainfall are validated against operational daily rain gauge analyses and summed radar rainfall estimates, with results updated on a daily basis. Validation results for 24 h precipitation forecasts from numerical weather prediction (NWP) models (a potential alternative source of rainfall information for users) are included for comparison.

There were three motivations for this project. First, algorithm developers benefit by being able to compare their estimates against the observed rainfall, as well as other algorithms, on a frequent basis. This helps them to discover systematic errors in their results, which is the first step to making improvements. Second, there are a myriad of estimates available on the web, as well as a variety of potential users of this information, and it was felt that users could benefit by a better understanding of the strengths and weaknesses of the satellite estimates. For unwary users the temptation is to assume that the estimates are perfectly accurate, when it is well known within the satellite precipitation community that the estimates contain significant systematic, as well as random, error.

Finally, as gridded rainfall estimates become increasingly available from NWP models, it is of scientific and practical interest to investigate when, where, and for which situations the satellite estimates are more accurate than NWP (indeed, can they be used to validate the NWP forecasts?), and visa versa. A recent study by Ebert et al. (2003) showed that the models had greatest accuracy for mid-latitude, large-scale, and wintertime rainfall, and poorer accuracy for tropical, convective, and summertime rainfall. No comparisons against satellite estimates were done in that study, but it is clear that the strengths and weaknesses of satellite estimates and NWP models are *complementary*.

2. OUTLINE OF THE VALIDATION / INTERCOMPARISON STUDY

While validation / intercomparison study is an ongoing project of the IPWG, an evaluation of the seasonal dependence of the algorithms' behaviour can begin after collecting at least one year of estimates. Estimates from several different regions of the globe are being processed, including Australia, the United States (Janowiak, 2004), and western Europe (Kidd, 2004). In addition, IPWG participants with reference data from Japan, Taiwan, South America, and West Africa have indicated a strong interest in participating in this study.

The time scale of primary interest to this study is daily rainfall, largely because the bulk of the rain gauge observations available for use in algorithm validation are 24 h accumulations. The Australian validation makes use of an operational daily rain gauge analysis at 0.25° resolution (Weymouth et al., 1999). The US validation has both a 24 h gauge analysis and the accumulated hourly Stage IV radar estimates (both 0.25°) as reference data. In western Europe only radar analyses (20 km and 50 km resolution) are available in a timely fashion for validation purposes.

The national gauge and radar analyses are less accurate than the high quality radar and rain gauge ground validation sites used in the GPCP Surface Reference Data Center or the TRMM

Ground Validation program. This means that the resulting error estimates will also not be as accurate as when "supersites" are used. However, a national scale validation can give a better "big picture" evaluation than is possible using only supersites.

The collected global satellite precipitation estimates are archived at the University of Maryland so that investigators can easily obtain data for future investigations.

At the time of writing, rainfall estimates from nineteen satellite algorithms are being evaluated in this study. These are all producing 24 h accumulated rainfall estimates on a daily basis in near-real time (within two days). Some of the algorithms are "operational" in the sense that they run continuously, output rainfall estimates to the web in graphical form, and provide the data in digital form to users via FTP. Other algorithms are "experimental", that is, results are not released to the public, or they are test versions of algorithm upgrades. Four NWP models are also included: three global models (resolution of 0.5° to 1.0°) and one regional mesoscale model (0.125° resolution). Each algorithm is validated its own spatial resolution or at the resolution of the gridded validation data, whichever is greater. An additional offline validation is performed at 1.0° resolution to enable fairer intercomparison with each other, with the coarser resolution NWP models, and with the GPCP 1-degree daily product.

A variety of validation approaches are used to measure different aspects of algorithm quality. To measure skill for rain occurrence, categorical statistics such as frequency bias, probability of detection (POD), false alarm ratio (FAR), equitable threat score (ETS), and Heidke Skill Score (HSS) are used, with results reported for a range of rain thresholds. (For definitions of these scores please refer to a textbook such as Wilks (1995) or the web page of the JWGV (2004).) To quantify errors in rain amount, we compute the mean error, mean absolute error, RMS error, and correlation coefficient. In addition, the object-oriented Contiguous Rain Area (CRA) method of Ebert and McBride (2000) is used to check the properties of rain systems in the Australian validation. The CRA method also decomposes the total error into components due to location errors (assumed negligible in this application), volume errors and pattern errors. The validation results are stratified by algorithm type (MW, IR, and MW-IR), region, season, and rain amount threshold, to further elucidate the nature of the errors.

3. VALIDATION RESULTS

The satellite rainfall validation results are updated daily and can be viewed on the web at

http://www.bom.gov.au/bmrc/wefor/staff/eee/SatRainVal/sat_val_austr.html (Australia)

http://www.cpc.ncep.noaa.gov/products/janowiak/us_web.shtml (US)

http://kermit.bham.ac.uk/~kidd/ipwg_eu/ipwg_eu.html (Europe)

The Australian web page includes validation of daily estimates, monthly and seasonal summaries of the daily results, validation of monthly estimates, and annual summaries of monthly results. The daily or monthly *results* give the performance for a particular spatial estimate time. The *summaries* aggregate the results to longer time scales and present the statistics as several diagnostic plots. A "user" page shows validation results for the operational algorithms and NWP models only, while the "developer" page adds the experimental algorithms and also provides a multi-algorithm map for visual comparison. This section describes the daily validation products found on the Australian web page.

The multi-algorithm map is a quick-look display of the rain gauge analysis and all of the satellite and NWP estimates. An example is shown in Figure 1. This display enables an easy comparison of the different products. It is seen, for example, that the MW products tend to be much less spatially

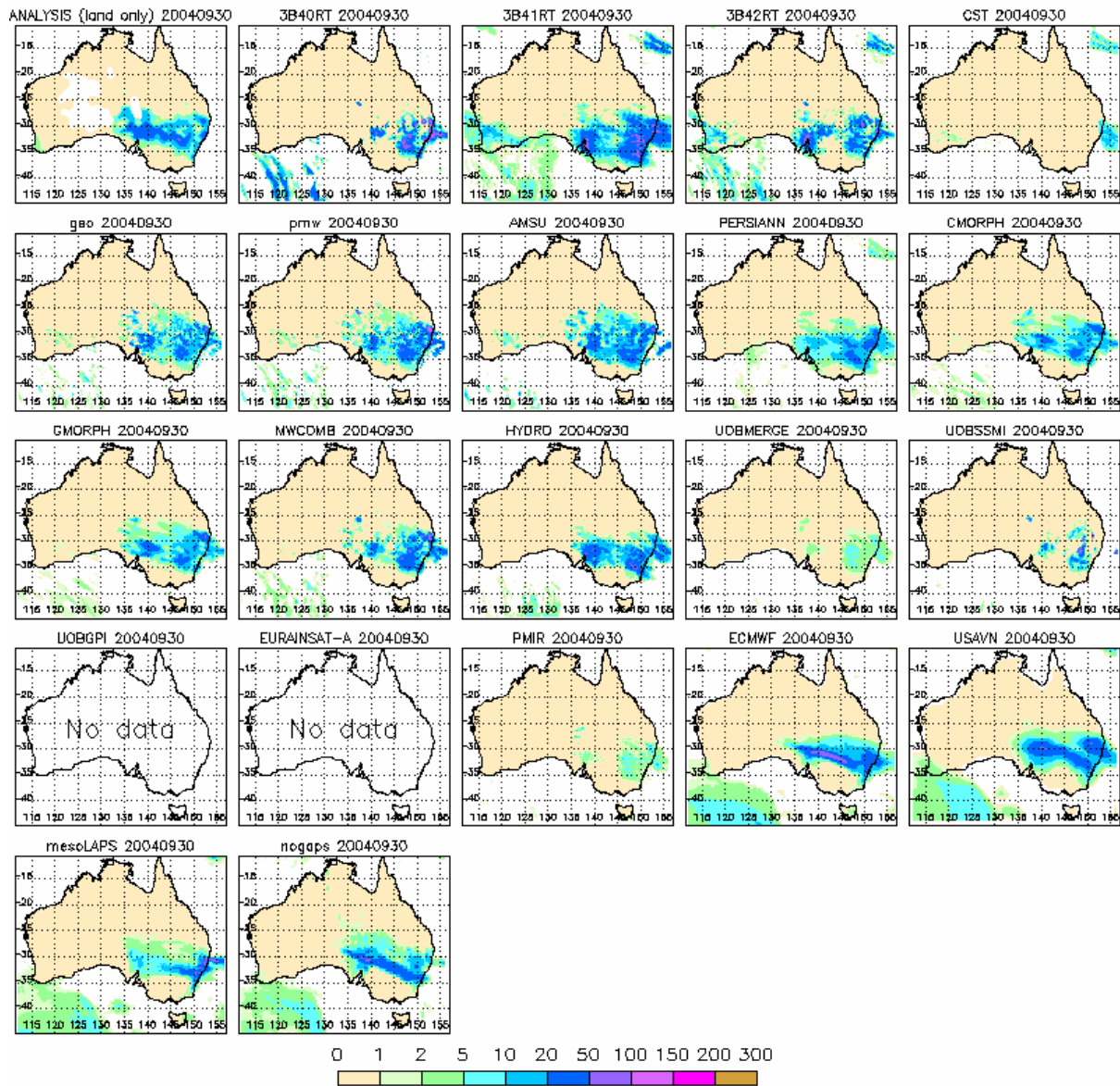


Fig 1. Multi-algorithm map for satellite and NWP rainfall estimates on 30 September 2004. The rain gauge analysis is shown in the upper left corner.

smooth than the IR and MW-IR products. We cannot overemphasize the value of frequent visual checking and intercomparison for helping to understand algorithm behaviour.

An example of a daily validation graphic is shown in Figure 2. On the left is the mapped satellite estimate and on the right is the operational daily gauge analysis for the same day. Note that there is a large region with no data in the gauge analysis; no attempt is made to validate satellite estimates in this region. A scatter plot (lower left) shows the direct correspondence between the estimates and the analysed rain amounts. The contingency table gives information on the correct prediction of rain occurrence, showing the frequency of hits, misses, false alarms, and correct rejections for a rain threshold of 1 mm d^{-1} . The statistics give the accuracy measures for the entire domain.

The full-domain statistics may include the effects of several unrelated rain systems. The CRA validation isolates the single largest rain system each day, defined as the contiguous rain area with the greatest rain volume in a maximum value image ($R_{ij} = \max[(R_{anal})_{ij}, (R_{sat})_{ij}]$), and validates the properties of the estimated rain entity against the properties of the analysed rain entity. In the example shown in Figure 3, the maximum rain rate was well estimated, but the volume of rain in

the system was only about ¼ of the observed amount. For this storm 90% of the total error was associated with errors in the fine scale structure, or pattern.

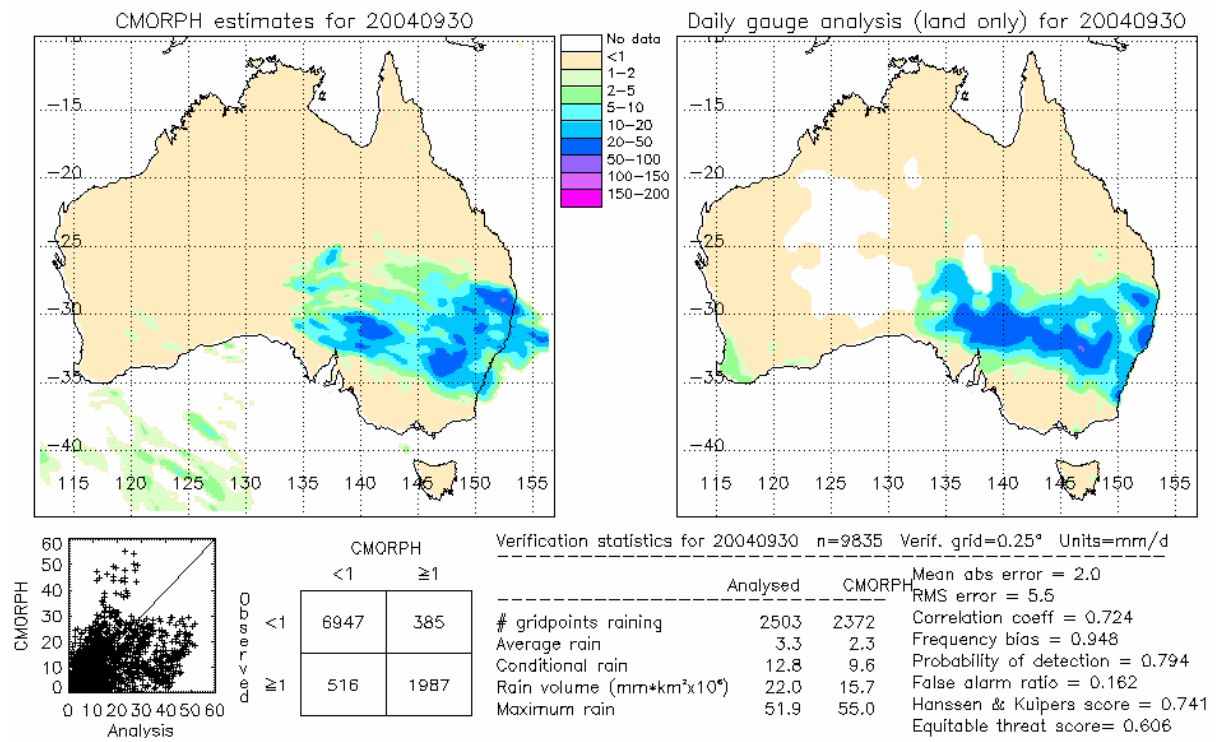


Fig 2. Validation of 24 h rainfall estimated by the CMORPH algorithm on 30 September 2004.

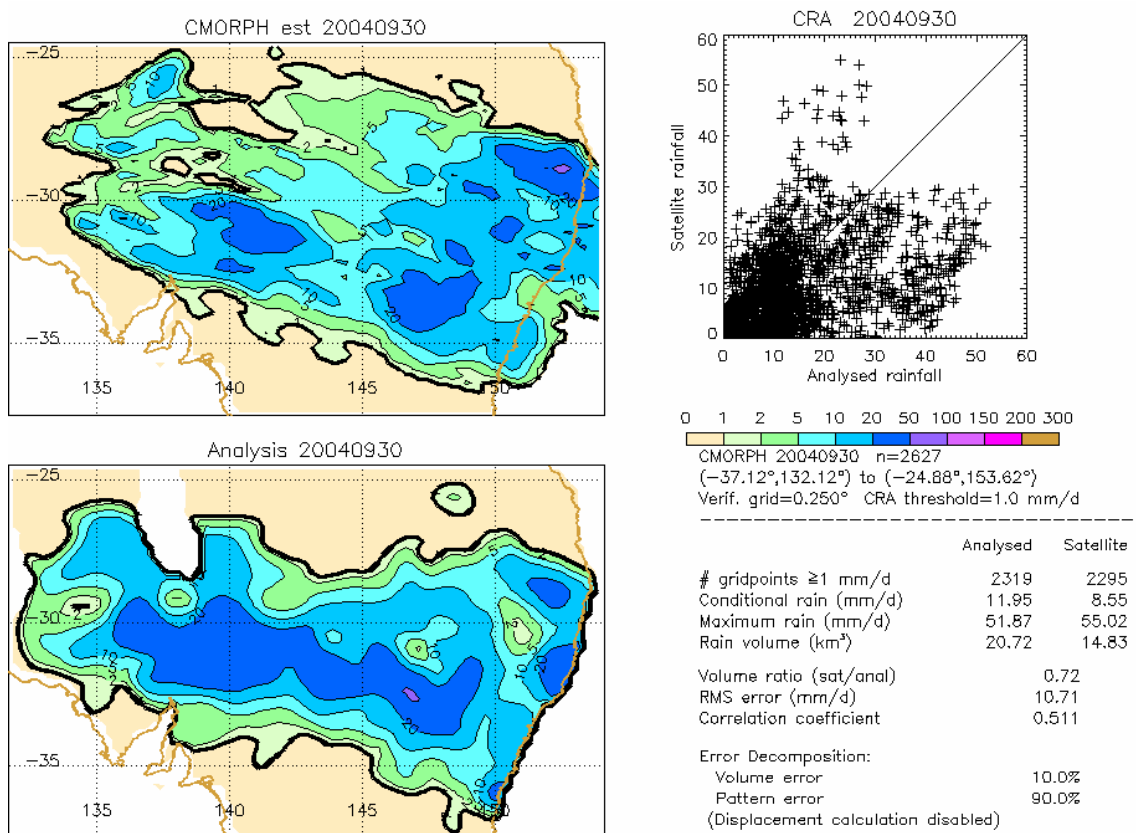


Fig 3. CRA validation of CMORPH estimated rainfall for the rain system in southeastern Australia on 30 September 2004.

Monthly and seasonal summaries aggregate the daily results over longer time periods to get a broader view of overall algorithm behaviour. At the moment these algorithms are grouped by provider; in the future they will also be grouped by algorithm type. Examples of seasonal summary statistics are the time series, scatter plots, tabulated statistics, binary (categorical) scores by threshold, and error magnitude as a function of rain range shown in Figures 4-8. By providing several different "views" of the errors it is possible to come to a better understanding of the algorithms' strengths and weaknesses.

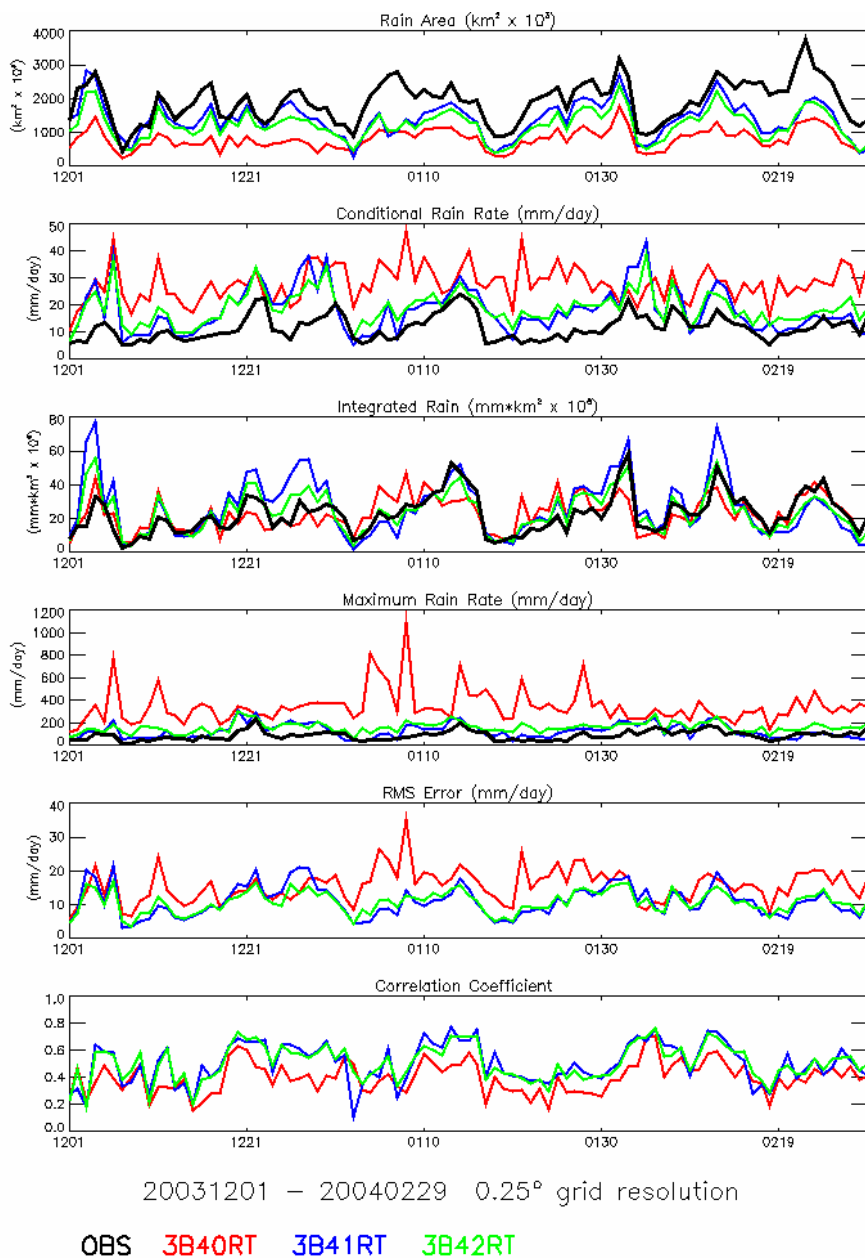


Fig 2. Time series of selected daily validation statistics for three GSFC algorithms during Dec. 2003 - Feb. 2004.

Some direct comparison of the satellite estimates with each other and with the NWP model forecasts on a 1.0° spatial grid has begun. Early results (Figure 9) suggest that in the tropics the satellite algorithms capture rain occurrence slightly better than the NWP models during summer but not during other seasons, while in mid-latitudes the NWP models outperform the satellite algorithms for rain occurrence during all seasons.

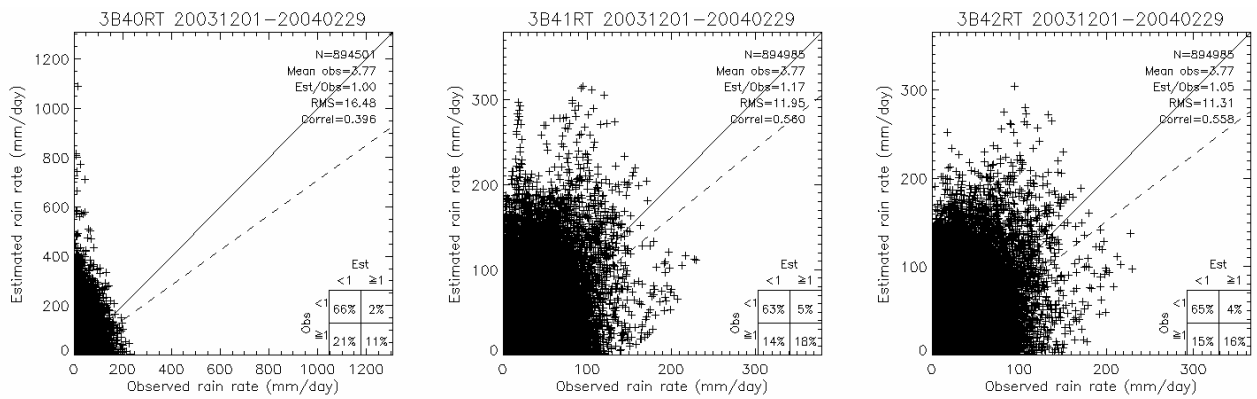


Fig 5. Scatter plots of estimated vs. observed daily rainfall for three GSFC algorithms during Dec. 2003 - Feb. 2004.

Comparative Statistics averaged for 20031201–20040229

	3B40RT	3B41RT	3B42RT
	AUST	AUST	AUST
OBSERVED			
Rain Area (km ² * 10 ³)	1922.	1924.	1924.
Avg Intensity (mm/d)	11.89	11.88	11.88
Rain Volume (mm*km ² * 10 ⁶)	22.85	22.86	22.86
Max Intensity (mm/d)	90.54	90.54	90.54
ESTIMATED			
Rain Area (km ² * 10 ³)	807.	1401.	1231.
Avg Intensity (mm/d)	28.33	19.15	19.53
Rain Volume (mm*km ² * 10 ⁶)	22.87	26.84	24.05
Max Intensity (mm/d)	346.39	127.45	162.33
Validation Statistics			
# Gridpoints	9829	9835	9835
Mean Abs Error (mm/d)	4.91	4.32	4.11
RMS Error (mm/d)	15.78	11.07	10.84
Correlation Coefficient	0.400	0.507	0.509
Bias Score	0.420	0.729	0.640
Probability of Detection	0.346	0.564	0.513
False Alarm Ratio	0.177	0.226	0.199
Critical Success Index	0.322	0.484	0.455
Hanssen & Kuipers Score	0.311	0.487	0.454
Equitable Threat Score	0.226	0.356	0.335

Fig 6. Table of validation statistics for three GSFC algorithms during Dec. 2003 - Feb. 2004.

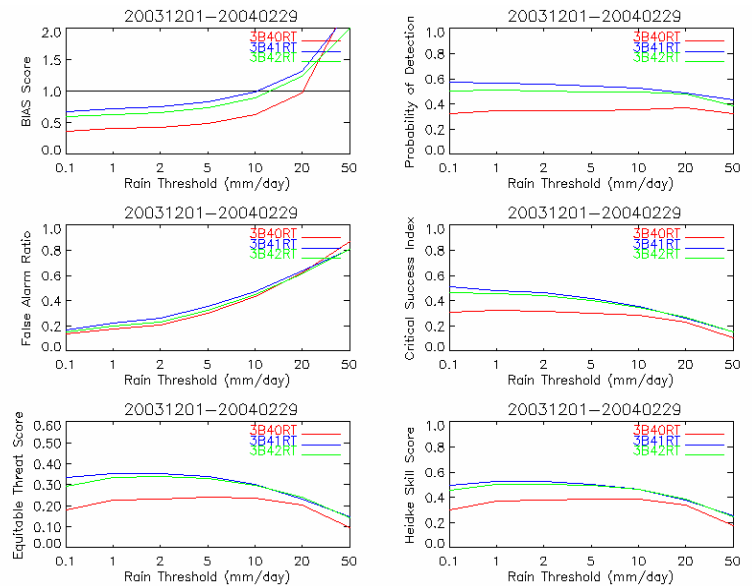


Fig 7. Categorical validation scores as a function of rain threshold for three GSFC algorithms during Dec. 2003 - Feb. 2004.

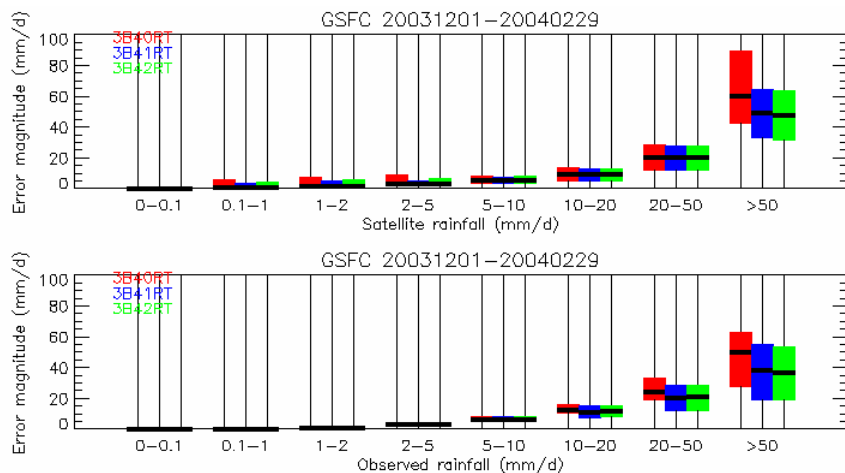


Fig 8. Mean absolute error as a function of estimated or observed rain amount, for three GSFC algorithms during Dec. 2003 - Feb. 2004. The boxes show the middle 50% of the distribution, the vertical lines show the full range of values, and the horizontal lines indicate the median.

It is perhaps unfair to lump all satellite algorithms within a type together, since some are expected to perform better than others. Figure 10 shows an example of the frequency bias for each MW-IR algorithm, plotted as a function of rain threshold. In the tropics all algorithms except PMIR and CST are biased high for heavy rainfall. In mid-latitudes all of the algorithms have a low bias (~50-70%) for lighter rainfall, while results are mixed for heavier rain.

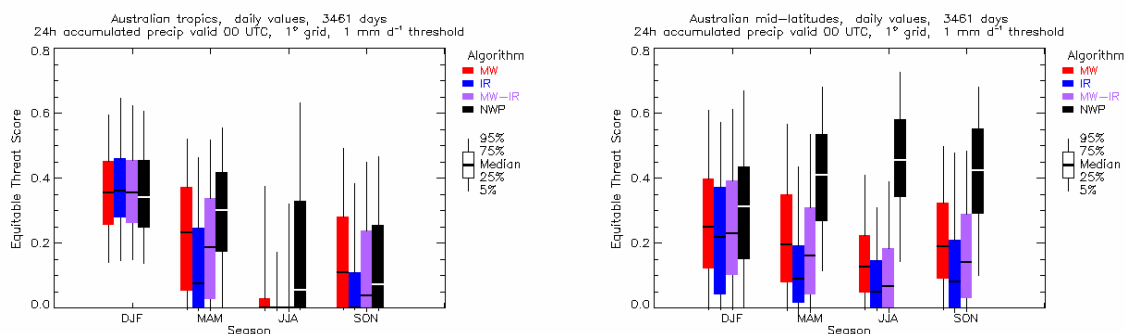


Fig 3. Equitable threat score for three classes of satellite algorithms and for NWP models, as a function of season, for the Australian tropics (left) and mid-latitudes (right) during Dec. 2002 – Sept. 2004.

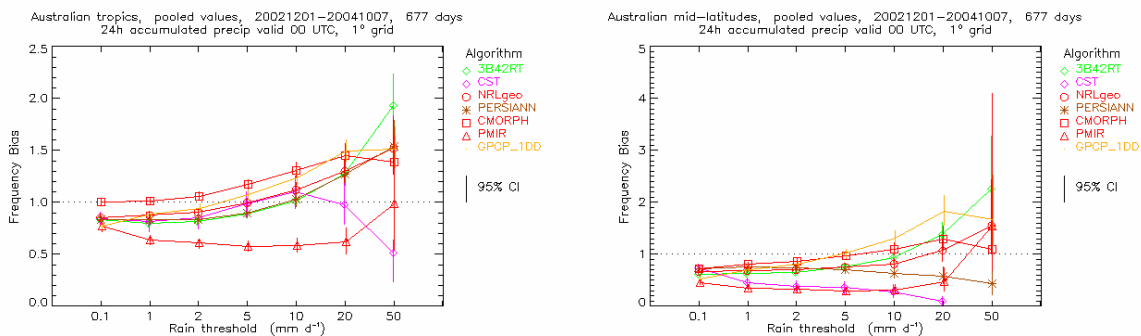


Fig 4. Frequency bias as a function of rain threshold for several MW-IR algorithms, for the Australian tropics (left) and mid-latitudes (right) during Dec. 2002 – Sept. 2004.

4. OUTLOOK

The IPWG algorithm validation / intercomparison has been producing results for more than a year, and will continue validating the algorithms for some time into the future. Feedback is sought from algorithm developers and precipitation data users on whether the validation web sites are giving the sort of information needed to improve, interpret, and appropriately use the rainfall estimates.

The hard work of interpreting the many results is just beginning. We hope that the validation will help clarify deficiencies in the algorithms and suggest ways in which they can be addressed. As various algorithm strategies are seen to work especially well, we anticipate that this will encourage a convergence of the satellite algorithms toward the most successful methodologies. This project provides algorithm developers with an opportunity to test new and improved algorithms and compare their output to other rainfall products. New "submissions" are welcome.

5. REFERENCES

Ebert, E.E., U. Damrath, W. Wergen and M.E. Baldwin, 2003: The WGENE assessment of short-term quantitative precipitation forecasts. *Bull. Amer. Met. Soc.*, **84**, 481-492.

- Ebert, E.E. and J.L. McBride, 2000: Verification of precipitation in weather systems: Determination of systematic errors. *J. Hydrology*, **239**, 179-202.
- Janowiak, J., 2004: Validation of satellite-derived rainfall estimates and numerical model forecasts of precipitation over the United States. *2nd Intl. Precipitation Working Group, 25-28 October 2004, Monterey, CA.*
- Joint Working Group on Verification (JWGV), 2004: *Forecast Verification: Issues, Methods, and FAQ.* http://www.bom.gov.au/bmrc/wefor/staff/eee/verif/verif_web_page.html.
- Kidd, C., 2004: Validation of satellite rainfall estimates over the mid-latitudes. *2nd Intl. Precipitation Working Group, 25-28 October 2004, Monterey, CA.*
- Weymouth, G., G.A. Mills, D. Jones, E.E. Ebert, and M.J. Manton, 1999: A continental-scale daily rainfall analysis system. *Aust. Met. Mag.*, **48**, 169-179.
- Wilks, D.S., 1995: *Statistical Methods in the Atmospheric Sciences. An Introduction.* Academic Press, San Diego, 467 pp.